(4)

AD-A269 612

Technical Report 1386
July 1992

A Revised Approach to Combining Linguistic and Probabilistic Information in Correlation

I. R. Goodman
Code 421

Technical Report 1386
July 1992

A Revised Approach to Combining Linguistic and Probabilistic Information in Correlation

I. R. Goodman
Code 421

NAVAL OCEAN SYSTEMS CENTER
San Diego, California 92152-5000

J. D. FONTANA, CAPT, USN
Commander

R. T. SHEARER
Technical Director

ADMINISTRATIVE INFORMATION

This report summarizes work performed during FY 1990 in the Command and Control Department (Code 40), Ashore Command and Intelligence Centers Division (Code 42), Advance Concepts and Development Branch (Code 421). The effort was supported by the Independent Research/Independent Exploratory Development (IR/IED) Program of the Naval Ocean Systems Center (NOSC), San Diego, California.

Released by
M. C. Mudurian, Head
Advanced Concepts
and Development Branch

Under authority of
J. A. Salzmann, Head
Ashore Command and
Intelligence Centers
Division

ACKNOWLEDGMENTS

The author wishes to express his gratitude to NOSC's Independent Exploratory Development Program (Code 0142) and Independent Research Program (Code 0141) for supporting this effort, as well as to his immediate line management at NOSC (Codes 421 and 42) for their overall general support and encouragement.

RT

SUMMARY

OBJECTIVE

The objective of this report is to present a general algorithm for performing data association or "correlation" for pairs of cumulative track data sets. Here, each pair represents potentially the same target or platform of interest; and the algorithm is required to (1) operate on both linguistic and stochastic information and (2) be feasible to implement. This means that in addition to the now well-established use of solely probabilistic information—usually in the form of geolocation and/or onboard radar information—linguistic or subjective attribute information is also to be used—such as in narrative descriptions of the form "appears to be a type X ship flying a dark flag and quite long (over 800 feet or so?), according to our intelligence reports."

A previous attempt in the same direction—though apparently reasonably successful as demonstrated through practical numerical experiments—indeed had the drawback of being replete with ad hoc components. The resulting algorithm, dubbed "PACT" (Possibilistic Approach to Correlation and Tracking), is fully documented in reference 1. The approach taken here, though preserving the main structure of PACT, aims at reducing its empirical and ad hoc aspects, as well as its relatively long running times.

RESULTS

A detailed presentation of the derivation and basic properties of the revised algorithms is given in this report, followed by tables and flowcharts that explicitly exhibit how to implement the algorithms. Unfortunately, at the time of the original writing of this report, only preliminary numerical verifications were carried out. Apropos to the task of final implementation, a group of four individuals was organized, lead by G. F. Kramer, Senior Analyst; and also comprised of this author; Dr. P. G. Calabrese, Senior National Research Council Fellow at NRaD; and Dr. C. J. Funk, Senior Analyst, Code 422. In turn, this has resulted in the NRaD Independent Exploratory Development supported TR, Application of the PACT Algorithm to Undersea Surveillance Data Fusion.

In conclusion, the structure of PACT has been modified to place it on a firmer mathematical basis. In addition, a new alternate, simplified approximation version has been obtained that appears promising in reducing the relatively long running time of the previous version of PACT (reference 1).

RECOMMENDATIONS

Further extensive numerical testing of these new versions of PACT should be performed to demonstrate both their eventual real-world feasibility and improvement upon the traditional use of only probabilistic information.

CONTENTS

FIGURES

TABLES

1. INTRODUCTION

For completeness, the basic problem of multitarget correlation and an outline of the PACT algorithms approach to this are presented here. For further details concerning the structure of the original PACT algorithm, its implementation, and numerical experiments and simulations involving PACT, see reference 1.

The typical correlation problem involves, say, $n(\geq 2)$ *track histories* that have not yet been equated; i.e., decided which truly represent distinct targets of interest or which actually represent the same target or targets of noninterest. The latter includes false alarms, whales, noncombatants, etc. Call each such track history $trk_\ell, \ell = 1, \ldots, n$, which consists of one or more pieces of observed or reported data that is assumed to be categorized by four indices: ℓ referring to the parent track trk_ℓ, itself; T, for time; S, for source, such as specific electronic sensor system or human intelligence sources; and A for attribute or type of data. The latter can refer to *stochastically oriented attributes*, including: A_1—geolocation, such as two-dimensional positions and velocity, as well as estimated covariance matrices or ellipses of errors associated with them; A_2—characteristics of parameters of onboard target radar systems such as pulse repetition interval (PRI), scan rate, and type of system; and A_3—characteristics of some particular nonradar sensor system onboard the target that is reasonably described in a *probabilistic* manner, among others. On the other hand, index A for data can also refer to *linguistically based* or *subjective data*, where the associated numerical descriptors are *possibility* or *fuzzy-set membership functions*. Examples of this can include the following: A_4—subjective classification, where the boundaries of the different classes are unclear and/or many overlaps can occur; A_5—narrative descriptions in natural language of target maneuvers or paths; and A_6—narratives of target characteristics, such as color of the flag, shape of ship, length; and so on.

Before further expounding upon, and analyzing fuzzy-set membership functions and their connections with probability, let us return to the main statement of the problem. Summarizing the basic situation, the following diagram (table 1-1) presents the scheme of information received.

Table 1-1. Scheme of information for various track histories.

trk_1	trk_2	. . .	trk_n
consists of data, each typically written as $y_{A,S,T,1}$	consists of data, each typically written as $y_{A,S,T,2}$	. .	consists of data, each typically written as $y_{A,S,T,n}$
where A, S, T run over appropriate values relevant to track$_1$	where A, S, T run over appropriate values relevant to track$_2$	. .	where A, S, T run over appropriate values relevant track$_n$

A = Attribute, S = Source, T = Time

Next, taking a simplified approach to correlation through sequential paired comparisons (but, see reference 2, page 13), for comments on the potential loss of accuracy by using this approach), consider each possible distinct pair, say $(trk_{\ell_1}, trk_{\ell_2})$, for $1 \leq \ell_1 < \ell_2 \leq n$, yielding a total of $\binom{n}{2} = \frac{n \cdot (n-1)}{2}$ such pairs.

Without loss of generality, denote trk_{ℓ_1} as simply trk_1 and trk_{ℓ_2} as trk_2, but keeping in mind the actual track numbers of subsequent processing. In addition, for completeness, from now on, for each fixed index A and ℓ, replace

$$\text{data set for } trk_\ell = \{y_{A,T,\ell} : (A,S,T) \text{ runs over the possible combination of values for } trk_\ell\} \tag{1.1}$$

by the single updated best possible value updated data set for trk_ℓ : $\{y_{A,To,\ell}\}$.

where T_o refers to the present time as a reference. When $A = A_1$, typically a Kalman filter or linear regression estimator can be used to produce $y_{A,To,\ell}$. Similarly, for other stochastic attributes such as A_2, A_3, here. When A is a linguistic-based attribute, the process of single data replacement-updating can be more complicated (or even simpler), depending upon the specific attribute in question. One could use a time gate or majority filtering technique or some transition Markov-like matrix approach in that case for the updating. For example,

$$\left.\begin{aligned} y_{A_6,T_{-3},3} &= \text{brown (flag color)}, \\ y_{A_6,T_{-2},3} &= \text{black}, \\ y_{A_6,T_{-1},3} &= \text{brown}, \end{aligned}\right\} \tag{1.2}$$

may simply distill to

$$y_{A_6,3} = \text{brown}, \tag{1.3}$$

omitting the present time index T_o.

In any case, a typical tableau of comparisons is shown in table 1-2.

Table 1-2. Tableau of data comparisons for a pair of typical track histories.

Attribute A	$y_{A_{i,1}}$ for trk_1 :	$y_{A_{i,2}}$ for trk_2 :
A_1	Mean Pos: 14°5'W, long 8°2'N lat 90% conf. ellipse: semimajor axis lngth: 0.3 nm semiminor axis lngth: 0.5 nm Angle of inclination: 16°	Mean Pos: 14°4'W, long 8°4'N, lat 90% conf. ellipse: semimajor axis lngth: 0.5 nm semiminor axis lngth: 0.8 nm Angle of inclination: -43°
A_4	A-II class	A-III class
A_6	brown flag	dark brown flag
A_7	rather long (over 800 feet)	longish (over 500 feet)

In addition to the paired comparison of information, as presented typically in table 1-2, one must consider the relative errors or reliability of the data so that observed/reputed/predicted $y_{A_{i,\ell}}$ can be compared numerically with any potential true value, say $z_{A_{i,\ell}}$, $\ell = 1, 2$, $i = 1, 2, 3, \ldots, m$. Furthermore, the relative weight of importance assigned to each attribute in the comparison process should also be ascertained in some sense. For example, how much more or less important is geolocation (A_1) matching than, say, flag color (A_6) matching?

In the past—as well documented in references 2 and 3—geolocation information matching was almost soiely emphasized, with relatively little emphasis upon other attributes, except for course-confirming or course-disconfirming evidence use. However, with the advent of AI techniques, it is becoming clearer that other attribute information can play critical roles in decision making in general, and correlation in particular. (For example, see reference 4 for employing fuzzy logic in decisions useful in expert systems.)

2. THE PACT APPROACH

In the philosophy of the PACT approach, the following three types of information are combined: observed data (suitably updated), reliability/errors of data, and relative weights of importance in the form of inference rules. In short, the PACT approach centers around a generalization—in order to include subjective as well as stochastic attributes—of a conditional form of the well-known "total probability theorem." (For example, see reference 5, pages 39 and 40):

$$P(x|y) = \sum_{z \epsilon D} P\ (x|y,z) \cdot P(z|y)\ , \tag{2.1}$$

where P is a given probability, x is a parameter of interest, y is observed data, and z is an *auxiliary variable* introduced so that $P(x|y,z)$ and $P(z|y)$ are known legitimate conditional probability functions of arguments x and z, respectively.

In a purely probabilistic context, the expansion in equation (2.1) completely utilizes the three types of information just noticed: observed data corresponds to variable y, reliability/errors of data correspond to the condition $P(z|y)$, and relative weights of importance correspond to the different values for the conditional $P(x|y,z)$. Indeed, the latter can be thought of as a model for the *inference rule* connecting y and z to X. Naturally, a different choice for $P(x|y,z)$—depending upon the meaning of auxiliary variable z—will produce a different weighting scheme.

Of course, it may be feasible to obtain directly, without recourse to the introduction of auxiliary variable z, the prior P(x) and the conditional form P(y|x)—especially when the regression model holds

$$y = g(x) + h(W) \tag{2.2}$$

for some known (measurement) function g and known objective function h, and where W is a random variable (or random vector) statistically independent of x, representing measurement/observation error, also with a known probability function.

In the preceding case, it follows that the conditional probability function for y given x is obtainable from equation (2.2) as

$$P(y|x) = P(W = h^{-1}(y - g(x)))\ . \tag{2.3}$$

In turn, Bayes' Theorem yields, in place of equation (2.1),

$$\begin{aligned} P(x|y) &= P(y|x) \cdot P(x)/P(y)\ , \\ P(y) &= \sum_{x \epsilon D_0} P(y|x) \cdot P(x)\ , \end{aligned} \tag{2.4}$$

letting D_0 be the (discrete) domain of values of A with $P(y|x)$ evaluated through equation (2.3).

Of course, the Bayes' Theorem form in equation (2.4) is only useful when $P(y|x)$ can be obtained directly. However, it is often the case—and certainly so for the main application here to correlation—that

$P(y|x)$ is *not* readily obtainable due to the great complexity relating x to y, and that the appropriate introduction of auxiliary variable z will supply the needed link.

Specifically, if x corresponds to correlation and y to jointly observed geolocational and radar parameters, by letting z be the corresponding *true* values in place of y, if $P(x|y,z) = P(x|z)$ and $P(z|y)$ are known, equation (2.1) is preferred to equation (2.4).

Another issue, mentioned earlier, is the desire to incorporate linguistic-based or subjective information with stochastic information. Often, in effect, such information involves a descriptor that does not represent an ordinary set, but rather a fuzzy set A determined by a corresponding membership function $\phi(A) : D_A \rightarrow [0, 1]$, where for any $x \in D_A$, $\phi(A)(x)$ is interpreted as "the degree to which x belongs to A." Examples of this include A = "tall," "long," "happy," "cold;" and where attribute domain of values D_A can be a fixed population of people or ships. On the other hand, one can in a sense, reverse the roles of A and D_A by choosing, e.g., A to represent "typical ship" and D_A to consist of a set of classifications $\{C_1, .., C_m\}$ so that the evaluation $\phi(A)(C_j)$ means the degree to which C_j belongs to A, i.e., to which a typical ship (prior) is of C_j type. (See, for example, Dubas and Pride [reference 6] for background information on fuzzy sets.)

Further details and analysis of fuzzy sets representing linguistic descriptions are given in the following chapter. In any case, the main thrust of PACT in utilizing both linguistic-based and probability information is, first, to extend the purely probabilistic expansion in equation (2.1) to a fuzzy-set form for each inference rule used, and second, to combine the resulting posterior descriptions of correlation for each inference rule used into an overall posterior description.

The chief motivation for the current revision of PACT is to reduce the empirical or ad hoc aspects relative to the choice of logical operators—especially the use of material implication in fuzzy-set form for modeling conditioning in inference rules and the combining of inference rules. In addition, a new upper-bound approximation for PACT will be presented, potentially greatly reducing PACT's relatively long running times. Before obtaining these results, a fresh discussion is presented motivating the encompassing use of fuzzy sets relative to information that is in part linguistic-based and, in part, stochastic.

3. FUZZY SETS AND PROBABILITY: PRELIMINARIES

Let $F: \mathbf{R} \to [0,1]$ be a given function. Then, recall from elementary probability theory that letting

$$D_F \overset{d}{=} \{x : x \in \mathbf{R} \ \ \& \ F(x+\Delta) - F(x) > 0 \ , \text{ for all } \Delta > 0\} \tag{3.1}$$

be the support of F, then F is a cumulative probability distribution (cdf) over D_F if F satisfies:

(i) F is nondecreasing continuous from the right.

(ii) $$\lim_{\substack{x \to -\infty, \\ x \in D_F}} F(x) = 0 \ , \quad \lim_{\substack{x \to +\infty, \\ x \in D_F}} F(x) = 1 \ . \tag{3.2}$$

Note that if D_F is finite, say,

$$D_F = \{x_1, .., x_n\} \text{ with } x_1 < x_2 < \ldots < x_n \ , \tag{3.3}$$

$x_j \in \mathbf{R}$, $j = 1, \ldots, n$, then defining $f : D_F \to [0, 1]$ by

$$f(x_j) \overset{d}{=} F(x_j) - F(x_{j-1}) > 0, \qquad j = 1, .. , n \ , \tag{3.4}$$

with the convention

$$x_o = -\infty \ , \text{ whence } F(x_o) = 0 \ , \tag{3.5}$$

implying

$$f(x_1) = F(x_1) \ , \tag{3.6}$$

so that

$$0 \le f(x_j) \le 1 \ , \quad j = 1, . . , n \ , \quad \sum_{j=1}^{n} f(x_j) = 1 \ , \tag{3.7}$$

establishing f as a probability function associated with F. As an interpretation: *One and only one outcome $x_j \in D_F$ can occur at a time, with probability $f(x_j)$.*

Returning to the general case for D_F: associated also with F is the *nested* (or level) *random set*

$$S(F,U) = F^{-1}[U,1] = \{ x : x\epsilon\mathbf{R} \ \& \ F(x) \geq U \} \tag{3.8}$$

where U is any random variable (r.v.) that is uniformly distributed over [0,1]. Note that one has the basic one-point coverage relation

$$\begin{aligned} \Pr\ (x \in S(F,U)) &= Pr\ (U \leq F(x)) \\ &= F(x)\ , \quad \text{all } x\epsilon\mathbf{R}\,. \end{aligned} \tag{3.9}$$

Indeed, making use of the basic properties of pseudoinverses denoted by $(\ )^{\dagger}$ (see, for example, Goodman and Ngugen [reference 7], pp. 121-124), one can show, slightly abusing notation, that

$$S(F,U) = [F^{\dagger}(U), c_F]\ , \tag{3.10}$$

actually a nested random interval with constant fixed right end point c_F , where

$$c_F \overset{d}{=} F^{-1}(1) \cap D_F \tag{3.11}$$

and random variable $F^{\dagger}(U)$ for its left end point, noting r.v.

$$\begin{aligned} F^{\dagger}(U) &= \inf F^{-1}(U,1] \\ &= \inf\{x{:}x\epsilon\mathbf{R} \ \ \& \ U \leq F(x)\}\ , \end{aligned} \tag{3.12}$$

has precisely cdf being F itself!

Conversely, if V is any r.v. over some $D_F \subseteq \mathbf{R}$, with cdf F, say, and $c_F \overset{d}{=} \sup\ (D_F)$, then the random interval $[V, c_F]$ can be naturally identified with F via the *one-point coverage relation*:

$$Pr(x\epsilon[V,c_F]) = Pr(V \leq x) = F(x)\ , \tag{3.13}$$

for all $x\epsilon D_F$.

One can also associate directly with the probability function f for F when D_F is finite, say as in equation (3.7), the random nested set

$$\begin{aligned} S(f,U) &= f^{-1}[U,1] \\ &= \{x : x\epsilon D_F \ \& \ f(x) \geq U\} \end{aligned} \tag{3.14}$$

Note, again, the one-point coverage relation

$$Pr(x\epsilon S(f,U)) = f(x),\ \text{all } x\epsilon D_F\ , \tag{3.15}$$

and most importantly, since f is a probability function (see equation 3.7),

$$1 = \sum_{x \epsilon D_F} Pr(x \epsilon S(f, U)) \ . \tag{3.16}$$

The constructions in equations (3.11) through (3.16) are all special cases of the relations between fuzzy sets and random sets pointed out originally in references 8 and 9; and later in reference 7, Chapters 3 and 4.

Fuzzy sets and their operators provide a way to represent, relatively simply, numerical uncertainties of possibly overlapping complex events (references 6 and 7). The following examples and analysis should provide a good basis for illustrating this point and relating fuzzy sets to probability concepts:

Suppose an adjectival or descriptor attribute A is given with an associated numerical function $\phi(A) : D_A \rightarrow [0, 1]$. (At times for convenience — when not confusing — we can simplify previous domain notation.) To begin with, A *could* represent the abstraction corresponding to any given cdf $F: \mathbf{R} \rightarrow [0, 1]$ with support D_F, so that $\phi(A) = F$. One could label A here as "F-ish" or simply "F," and so on. $\phi(A)$ frequently does *not* behave like a cdf or probability function. Indeed, if A is first given as a linguistic descriptor or adjective relative to some appropriate domain D_A and then a suitable associated numerical-valued function $\phi(A) : D_A \rightarrow [0, 1]$ is sought for the interpretation, for each $x \epsilon D_A$:

$$\begin{aligned} \phi(A)(x) &= \text{"degree to which x has property A"} \\ &= \text{"membership level of x in fuzzy set A"} \\ &= \text{"possibility of A occurring at x"} \end{aligned} \tag{3.17}$$

Typical examples include the descriptors:

$$A_1 = \text{happy}, \quad D_{A_1} = \text{adult males living in Denmark} \tag{3.18}$$

$$A_2 = \text{long}, \quad D_{A_2} = [100', 1500'] \tag{3.19}$$

representing possible ship lengths,

$$A_3 = \text{dark}, \quad D_{A_3} = \text{spectrum of colors} \tag{3.20}$$

$$A_4 = \text{close}, \quad D_{A_4} = \mathbf{R}^{+} \tag{3.21}$$

representing distance between targets,

$$A_5 = \text{typical prior target}, \quad D_{A_5} = \{C_1, \dots, C_{10}\} \ , \tag{3.22}$$

where C_j represents the j^{th} class of ship, where, in general, the C_j's are overlapping. Note the effective reversal in A_5 where the domain values play the role of the descriptor and the attribute is a relatively neutral typical population element.

In a similar vein, one can make a fuzzy set interpretation of cdf's and pf's, noting again the role of the descriptors being determined by the domain values, rather than by given prior attritubes.

Even if in any natural interpretation of an attribute A, $\phi(A) : D_A \to [0, 1]$ is nondecreasing—and hence formally similar to a cdf—the interpretation of $\phi(A_2)(x)$ can be quite different than if $\phi(A)$ represented a cdf. For example, $\phi(A_2)$ has the form of a cdf, and, hence, relative large values of $x \epsilon D_A$, yield large values for $\phi(A_2)(x)$, as should be, considering the meaning of "long." On the other hand, one can question how useful is the attribute A_6 defined to correspond to cdf $\phi(A_6) : \mathbf{R}^2 \to [0, 1]$ to describe a target's location. Certainly, for $x \epsilon \mathbf{R}^2$ large positive in both components, $\phi(A_6)(x)$ is close to 1, but x can be quite far from the mean of the distribution. In this case, the local attribute, i.e., the discretized probability function corresponding to $\phi(A_6)$, is a more meaningful numerical measure.

4. FUZZY SETS AND PROBABILITY: GENERAL CASE

In any case, one is led to the following general random-set interpretation for fuzzy sets extending the material just covered in Chapter 3.

For any given attribute A with associated *possibility* or *fuzzy set membership function* $\phi(A) : D_A \rightarrow [0,1]$, define the *associated* (unique) *nested random set* $S(A,U) \subseteq D_A$ analogous to equation (3.8) or equation (3.14):

$$\begin{aligned} S(A,U) &= \phi(A)^{-1}[U,1] \\ &= \{x : x \in D_A \text{ and } U \leq \phi(A)(x)\}, \end{aligned} \tag{4.1}$$

U any r.v. uniform $[0,1]$ distributed.

It then follows immediately that for all $x \in D_A$, the one-point coverage relation holds

$$\begin{aligned} \Pr(x \in S(A,U)) &= \Pr(U \leq \phi(A)(x)) \\ &= \phi(A)(x) \ . \end{aligned} \tag{4.2}$$

When again D_A is finite, say $\{x_1, \ldots, x_n\}$, then equation (4.2) can be interpreted also as identifying the possibility of x occurring relative to A as the probability of the *filter class* or *class of interacting subsets* of D_A relative to x, $_x$, occurring, where

$${}_x \overset{d}{=} \{a : x \in a \subseteq D_A\} \tag{4.3}$$

Thus,

$$\begin{aligned} \phi(A)(x) &= \Pr(S(A,U) \in {}_x) \\ &= \sum_{a \in {}_x} \Pr(S(A,U) = a) \\ &= \underset{a \in {}_x}{Pr}(\text{Or } (a \text{ occurs})) \ , \end{aligned} \tag{4.4}$$

summarizing the fundamental relation between attribute possibility outcomes, or equivalently fuzzy set membership function levels, *and probabilities of interactive sets.*

Note, also, that any typical outcome $S(A,U) \subseteq D_A$ can be thought of as all elements of $S(A,U)$ occurring simultaneously.

Note the following three special cases for $S(A,U)$, depending upon the structure of $\phi(A) : D_A \rightarrow [0,1]$, slightly abusing notation (D_A need not be finite):

(i) $\phi(A)$ nondecreasing (but not necessarily a cdf).

$$S(A,U) = [\phi(A)^{\dagger}(U), c_A] \ , \tag{4.5}$$

where c_A is determined by the relation

$$\sup_{x \in D_A} \phi(A)(x) = \phi(A)(c_A) \ . \tag{4.6}$$

(ii) $\phi(A)$ nonincreasing.

Defining A' also over D_A via

$$\phi(A') \overset{d}{=} 1-\phi(A)$$
$$S(A,U) = D_A - S(A', 1-U) \tag{4.7}$$

$$= [d_A, \phi(A')^{\dagger}(1-U)]$$
$$= [d_A, \phi(A)^{\dagger}(U)] \text{ ,} \tag{4.8}$$

where d_A is determined by the relation

$$\inf_{x \in D_A} \phi(A)(x) = \phi(A)(d_A) \text{ .} \tag{4.9}$$

(iii) $D_A = \{x_1, \ldots, x_n\}$

$$S(A,U) = \{x_j : x_j \in D_A \ \& \ \phi(A)(x_j) \geq U\} \text{ ,} \tag{4.10}$$

noting if

$$0 \leq \phi(A)(x_j) < \ldots < \phi(A)(x_n) \leq 1 \text{ ,} \tag{4.11}$$

then

$$Pr(S(A,U) = \{x_j, x_{j+1}, \ldots, x_n\}) = \phi(A)(x_j) - \phi(A)(x_{j-1}) \text{ ,} \tag{4.12}$$

for $j = 1, \ldots, n$, with the convention – using a fictitious x_0 and x_{n+1} –

$$\phi(A)(x_0) = 0, \quad \phi(A)(x_{n+1}) = 1 \text{ ,} \tag{4.13}$$

so that

$$\left.\begin{array}{l} Pr(S(A,U) = \emptyset) = 1 - \phi(A)(x_n), \\ Pr(S(A,U) = \{x_1, \ldots, x_n\}) = \phi(A)(x_1) \end{array}\right\} \tag{4.14}$$

Next, note that there can be many distinct random-set representations in general of any given fuzzy set membership function. For example, if $\phi(A) : D_A \to [0, 1]$ is such that $D_A = \{x_1, \ldots, x_n\}$, then the random set $T(A) \subseteq D_A$ also represents $\phi(A)$ where for any subset $a \subseteq D_A$

$$Pr(T(A) = a) \overset{d}{=} \prod_{x \in a} \phi(A)(x) \cdot \prod_{x \in D_A - a} (1 - \phi(A)(x)) \tag{4.15}$$

$T(A)$ is the maximal entropy representative, while $S(A, U)$ has minimal entropy-like properties—in any case, very different in form (see, for example, reference 7, especially section 5A). Conversely, *any* random-subset $S \subseteq D_A$, say, determines a unique one-point coverage equivalent fuzzy set membership function, namely

$$\phi(A)(x) = Pr(x \in S), \qquad x \in D_A, \tag{4.16}$$

Thus, with the plethora of possible (one-point coverage) random-set representations for a given fuzzy set, which one(s) should we choose and how does this choice relate this to established fuzzy set logical operators such as Zadeh's proposed min—or product at times—for intersection, max—or probability sum (as follows) for union, and $1-(\ )$ for complement, etc.?

In response to the preceding discussion, note that the nested random-set representation for attributes A referring to cdf's reduce nicely to the natural fixed right-hand constant, left-hand r.v. internal case as in equations (3.8) through (3.13). For the other extreme cases, such as A corresponding to $\phi(A)$ being a nonincreasing membership function or to a legitimate probability function, also naturally corresponding forms held for the nested random-set representations. (Again, see equations (4.7) through (4.9) and (3.4) through (3.6).

As a consequence of the above, as well as for ease of interpreting logical operators over fuzzy sets through ordinary corresponding ones—as will be seen in the next chapter—the nested random-set representations will be considered from now on as the natural probabilistic interpretation of fuzzy sets.

5. FUZZY SETS AND PROBABILITY: OPERATOR RELATIONS

Taking the natural nested random set representation of fuzzy sets (see also Goodman [reference 10] for further justifications in terms of flou sets and extensions of the Negoita-Ralescu representations), consider the following development, based upon the concept of *copulas*, i.e., cdf's for n r.v.'s $U_1, \ldots, U_n$ each distributed unif–[0,1] (see reference 7, section 2.3.6, or Schweizer and Sklar [reference 11] for background).

Roughly speaking, copulas can be considered to be natural generalizations of ordinary two-valued conjunction. Similar comments are valid for cocopulas—i.e., DeMorgan and related transforms of copulas representing generalizations of ordinary disjunction. Examples of copulas and cocopulas are (min, max), (prod, probsum) where

$$\text{probsum}\ (t_1, \ldots, t_n) \overset{d}{=} 1 - \prod_{j=1}^{n} (1 - t_j), \quad t_j \in [0,1], \quad j=1, \ldots, n. \tag{5.1}$$

Another copula, cocopula pair, among an infinitude of possibilities, for $n=2$, is (maxbnd, minbnd) where

$$\text{maxbnd}\ (t_1, t_2) = \max\ (t_1 + t_2 - 1, 0)\ , \tag{5.2}$$

$$\text{minbnd}\ (t_1, t_2) = \min\ (t_1 + t_2, 1)\ , \tag{5.3}$$

for all $t_j \in [0, 1]$.

Then, denoting generically cop: $[0,1]^n \to [0,1]$ for the *(joint-n)* cdf of $U_1, \ldots, U_n$, for any attributes A_j with domain D_{A_j}, $j = 1, \ldots, n$, define *conjunction*, $A_1 \mathbin{\&} \cdots \mathbin{\&} A_n$, *disjunction* $A_1 \vee \cdots \vee A_n$ and *negation* A'_j through the corresponding membership functions, where $\phi(A_1 \mathbin{\&} \cdots \mathbin{\&} A_n): D_{A_1} \times \cdots \times D_{A_n} \to [0,1]$ is given by

$$\phi(A_1 \mathbin{\&} \cdots \mathbin{\&} A_n)(x_1, \ldots, x_n) = \text{cop}(\phi(A_1)(x_1), \ldots, \phi(A_n)(x_n))\ , \tag{5.4}$$

and where $\phi(A_1 \vee \cdots \vee A_n) : D_{A_1} \times \cdots \times D_{A_n} \to [0,1]$ is given by

$$\begin{aligned}\phi(A_1 \vee \cdots \vee A_n)(x_1, \ldots, x_n) &= \text{cocop}(\phi(A_1)(x_1), \ldots, \phi(A_n(x_n)) \\ &\overset{d}{=} \sum_{\emptyset \neq K \subseteq \{1, \ldots, n\}} (-1)^{\text{card}(K)} \cdot \text{cop}((\phi(A_j)(x_j)); j \in K)\end{aligned} \tag{5.5}$$

and

$$\phi(A'_j(x_j)) \overset{d}{=} 1 - \phi(A_j)(x_j)\ , \tag{5.6}$$

for all $x_j \in D_{A_j}$ $\quad j = 1, \ldots, n$.

Sklar's Theorem (reference 11) is compatible with the result that $\phi(A_1 \mathbin{\&} \cdots \mathbin{\&} A_n)$ is a legitimate cdf over $D_{A_1} \times \cdots \times D_{A_n}$ if each $\phi(A_j)$ is a cdf over $D_{A_j} \in \mathbf{R}$, $j = 1, \ldots, n$. Conversely, any cdf F over $D_{A_1} \times \cdots \times D_{A_n} \subseteq \mathbf{R}^n$ can be expressed as

$$F = F_1 \& \cdot \cdot \& F_n \overset{d}{=} \mathrm{cop}_\circ(F_1, \ldots, F_n) \ , \tag{5.7}$$

where F_j is the j^{th} marginal cdf of F, $j=1, \ldots, n$.

In addition, the basic homomorphic tie-in between $\&$, $\vee$, $(\ )'$ relative to all fuzzy sets and ordinary $\cap, \cup, (\ )'$ relative to nested random sets can be established:

For all $x_j \epsilon D_{A_j}, j = 1, \ldots, n$, by the definition of cop

$$\phi(A_1 \ \& \cdot \cdot \ \& \ A_n)(x_1, \ldots, x_n) = Pr((U_1 \le \phi(A_1)(x_1)) \ \& \cdot \cdot \ \& \ (U_n \le \phi(A_n)(x_1))) \tag{5.8}$$

$$= Pr((x_1 \epsilon S(A_1, U_1)) \ \& \cdot \cdot \ \& \ (x_n \ \epsilon \ S(A_n, U_n)))$$

Hence, when $D_{A_1} = \cdot \cdot = D_{A_n} = D$, for all $x \epsilon D$ (equation 5.8) implies

$$\begin{aligned} \phi(A_1 \cap \cdot \cdot \cap A_n)(x) &\overset{d}{=} \phi(A_1 \ \& \cdot \cdot \ \& \ A_n)(x, \ldots, x) \\ &= Pr(x \epsilon S(A_1, U_1) \cap \cdot \cdot \cap S(A_n, U_n)) \ , \end{aligned} \tag{5.9}$$

For all $x_j \ \epsilon \ D_{A_j}, j = 1, \ldots, n$, by reference 11, the definition of cop and the basic Poincare alternate sum expansion of probabilities,

$$\begin{aligned} \phi(A_1 \vee \cdot \cdot \vee A_n)(x_1, \ldots, x_n) &= \sum_{\phi \neq K \subseteq \{1, \ldots, n\}} (-1)^{\mathrm{card}(K)+1} \, Pr(\underset{j \in K}{\&} (U_j \le \phi(A_j)(x_j))) \\ &= Pr(\bigvee_{j=1}^{n} (U_j \le \phi(A_j)(x_j))) \\ &= Pr((x_1 \epsilon S(A_1, U_1)) \vee \cdot \cdot \vee (x_n \epsilon S(A_n, U_n))) \end{aligned} \tag{5.10}$$

Hence, when $D_{A_1} = \cdot \cdot = D_{A_n} = D$, for all $x \epsilon D$ (equation 5.10) implies

$$\begin{aligned} \phi(A_1 \cup \cdot \cdot \cup A_n)(x) &\overset{d}{=} \phi(A_1 \vee \cdot \cdot \vee A_n)(x, \ldots, x) \\ &= Pr(x \epsilon S(A_1, U_1) \cup \cdot \cdot \cup S(A_n, U_n)) \end{aligned} \tag{5.11}$$

Finally, it is obvious that for all $x_j \ \epsilon \ D_{A_j}$,

$$\begin{aligned} \phi(A_j'(x_j)) &= 1 - Pr(x_j \epsilon S(A_j, U_j)) \\ &= Pr(x_j \epsilon D_{A_j} - S(A_j, U_j)) \\ &= Pr(x_j \epsilon S(A_j', 1 - U_j)) \ . \end{aligned} \tag{5.12}$$

Equations (5.4) through (5.12), together with the previous justification for the use of unconditional fuzzy sets, logical operators in the form of copulas, cocopulas and 1-( ), model linguistic-based attributes and other logical connections, compatible with probability theory.

6. CONDITIONAL FUZZY SETS

Finally, consider the modeling of *conditional attributes*. Reference 10, section 9, shows that if A_1, A_2 are two given attributes, with any corresponding domains D_{A_1}, D_{A_2}, then the conditional attribute $(A_1|A_2)$—read "A_1 given A_2"—should be characterized for fixed copula, through the membership function $\phi(A_1|A_2) : D_{A_1} \times D_{A_2} \rightarrow [0,1]$, where for any $x_j \in D_{A_j}$, $j = 1, 2$,

$$\phi(A_1|A_2)(x_1, x_2) = \begin{cases} \phi(A_1 \& A_2)(x_1, x_2)/\phi(A_2)(x_2), \\ \qquad \text{provided } \phi(A_2)(x_2) > 0 \ , \\ [0,1] \\ \qquad \text{provided } \phi(A_2)(x_2) = 0 \ , \end{cases} \tag{6.1}$$

where $\phi(A_1 \& A_2)$ is given in equation (5.8) for n=2. Clearly, equation (6.1) reduces to a version of ordinary conditional cdf's when $\phi(A_1)$ and $\phi(A_2)$ are cdf's.

In turn, one can develop an entire calculus of logical operations extending $\&, v, (\ )'$—and hence $\cap, \cup, -$ as well—from the unconditional case to the more general conditional one. This is based on the calculus of conditional logical operations relative to nonfuzzy or ordinary *conditional events*. For background on conditional event algebra, see, for example, the recent summaries, reference 10, section 6; reference 12, or the forthcoming monograph, reference 13. In brief, if R is a Boolean algebra of ordinary sets or equivalently unconditioned events or propositions, then for all a, b, c, $d \in$ R, the following extensions of the ordinary Boolean logical operation $\&$ (or $\cap$), $\vee$ (or $\cap$), and$(\ )'$(or $-$) hold:

$$\left.\begin{aligned} (a|b) \ \& \ (c|d) &= (abcd|a'b \vee c'd \vee bd) \ , \\ (a|b) \vee (c|d) &= (ab \vee cd|ab \vee cd \vee bd) \ , \\ (a|b)' &= (a'|b) \ , \end{aligned}\right\} \tag{6.2}$$

where for shorthand

$$\left.\begin{aligned} abcd &= a \ \& \ b \ \& \ c \ \& \ d \ , \\ a'b &= a' \ \& \ b \ , \end{aligned}\right\} \tag{6.3}$$

and so on.

In turn, as developed in reference 10, section 9, one can use equation (6.2) to determine in a natural way the further extensions to conditional fuzzy sets.

This is accomplished through use of the nested random-set representation of the conditional fuzzy sets and their application of equation (6.2). For example, for any attributes A_j with domains D_{A_j}, $x_j \in D_{A_j}$ j=1,2,3,4, with a fixed choice of copula cop, letting

$$\alpha_j(x_j) \overset{d}{=} (x_j \epsilon S(A_j, U_j)) = (U_j \leq \phi(A_j)(x_j))$$

$$= U_j^{-1}[0, \phi(A_j)(x_j)] \; \epsilon R \text{ (boolean)} ,$$

j=1,2,3,4, one has

$$\phi((A_1|A_2) \vee (A_3|A_4))(x_1, x_2, x_3, x_4)$$

$$= Pr((\alpha_1(x_1)|\alpha_2(x_2)) \vee (\alpha_3(x_3)|\alpha_4(x_4))) \tag{6.4}$$

$$= Pr(\alpha_o|\beta_o) ,$$

$$\alpha_o \overset{d}{=} \alpha_1(x_1)\alpha_2(x_2) \vee \alpha_3(x_3)\alpha_4(x_4) ,$$

$$\beta_o \overset{d}{=} \alpha_o \vee (\alpha_2(x_2)\, \alpha_4(x_4)) . \tag{6.5}$$

In turn, equation (6.5) can be used to evaluate equation (6.4) fully. (Again, see reference 10, section 9, for further details.) Of course, one could utilize in place of the above development, a more simplified approach, where, for example, $\vee$ as in equations (6.4) and (6.5) is approached the same formally as in the unconditional case; for example, equation (5.5).

Finally, note that based on equation (6.1), one can show (see reference 10, equations (8.42) and (8.43)) that if cop is associative and distributive and $h : Z \to [0, 1]$ is any auxiliary function so that, using $\vee$ to indicate max or sup,

$$\max_{z \epsilon Z} \; h(z) = 1 , \tag{6.6}$$

Then, for all $x \epsilon X$, $y \epsilon Y$, the analog of equation (2.1) holds, that we call the *PACT expansion*:

$$(f(x)|g(y)) = \max_{z \epsilon Z} \; (f(x)|\mathfrak{g}(y,z)) \cdot (h(z)|g(y)) \tag{6.7}$$

where

$$\mathfrak{g}(y,z) \overset{d}{=} \text{cop}(h(z), g(y)) , \tag{6.8}$$

for all $y \epsilon Y$, $z \epsilon Z$

7. DEVELOPMENT OF MODIFIED PACT: MODIFIERS AND ATTRIBUTE-MATCHING FUNCTIONS

With all of the preliminaries in place, we can establish the new modified PACT algorithm. In the former version, a number of ad hoc assumptions were made, including:

(1) combining all inference rules beforehand into one joint rule using an arbitrary choice for conjunction

(2) identification of conditioning with a fuzzy-set version of material implication

(3) ad hoc forms for matching functions used to construct the inference rules.

In addition, PACT also had the drawback of requiring excessively long-running times due to the required disjunctive iteration of variables running over the auxiliary attribute domains. (Again, see reference 13 for further details.)

In brief, the modified version of PACT here addresses (1) by using only one inference rule $(\mathfrak{f} \mid \mathfrak{g})$ in equation (6.7) at a time, followed by combining the outputs. In (2), conditioning is assumed to be of the above form $(\mathfrak{f} \mid \mathfrak{g})$, rather than in material implication form. However, later, when an abbreviated version of PACT is sought, conditioning is replaced by a more expedient form. For (3), a matching function is considered simply a joint fuzzy-set membership function of the same marginal attribute connected by a copula. Finally, a simplified substitution and approximation relative to the PACT expansion is shown to lead to a greatly reduced running-time version of PACT.

Apropos to the fuzzy-set version of equation (1):

Let variable $x \epsilon [0, 1]$ represent the true level of overall *track association* or "*correlation*" of two given apparent targets of interest that may or may not actually represent the same target. Let attribute A_o correspond to correlation so that $\phi(A_o) : [0, 1] \to [0, 1]$ represents its membership function, which may well be simply the identity $id_{[0,1]}$.

Introduce *auxiliary attributes* and variables that provide potential information about A_o conditioned on any observed attribute values: $A_1, A_2, \ldots, A_m$. For example, one could have:

$$A_1 = \text{geo}_2 \text{ (two–dimensional locations)} \tag{7.1}$$

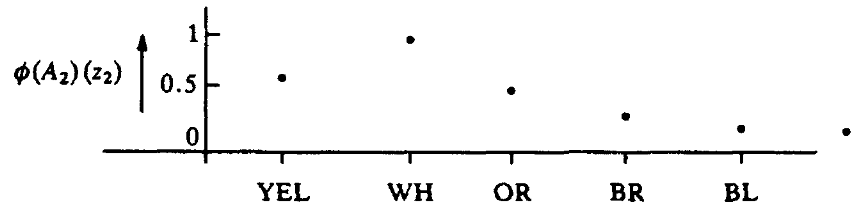
with $\phi(A_1) : D_{A_1} \to [0, 1]$, where $D_{A_1} \subseteq \mathbf{R}^2$ is some finite domain with $\phi(A_1)$ a suitable discretized-truncated version of Gaussian distribution $N_2(\mu_1, \Lambda_1), \mu_1 \epsilon \mathbf{R}^2$ mean and Λ_1, 2 by 2 positive definite covariance matrix of error, so that typically

$$\phi(A_1)(z_1) = \lambda_{11} exp - \lambda_{12}(z_1 - \mu_1)^T \Lambda^{-1}(z_1 - \mu_1), \qquad z_1 \epsilon D_{A_1} \tag{7.2}$$

for some positive scaling factors $\lambda_{11}, \lambda_{12}$; $\lambda_{11} \leq 1$. Or, one could have

$$A_2 = \text{color (dominant flag color of enemy)} \tag{7.3}$$

with associated membership function given numerically as



(7.4)

In general, $\phi(A_i) : D_{A_i} \rightarrow [0, 1]$ is assumed known from prior information with D_i some domain of values for attribute A_i, i=1, ..., m.

Next, suppose a range of modifiers has been established with associated membership functions, so that modifier $_j$, mod$_j$ has associated membership function $\phi(\text{mod}_j) : [0, 1] \rightarrow [0, 1]$. A simple collection of such modifiers would correspond to the linguistic and numerical interpretation as follows:

$$\begin{array}{lllllll}
\text{very high} & \leftrightarrow & \text{mod}_1 & \leftrightarrow & \phi(\text{mod}_1)(t) = t^{\alpha_1}, & t \in [0, 1] & \\
\text{high} & \leftrightarrow & \text{mod}_2 & \leftrightarrow & \phi(\text{mod}_2)(t) = t^{\alpha_2}, & t \in [0, 1] & \\
\text{moderately high} & \leftrightarrow & \text{mod}_3 & \leftrightarrow & \phi(\text{mod}_3)(t) = t^{\alpha_3}, & t \in [0, 1] & \\
\text{medium} & \leftrightarrow & \text{mod}_4 & \leftrightarrow & \phi(\text{mod}_4)(t) = t^{\alpha_4}, & t \in [0, 1] & (7.5) \\
\text{low medium} & \leftrightarrow & \text{mod}_5 & \leftrightarrow & \phi(\text{mod}_5)(t) = t^{\alpha_5}, & t \in [0, 1] & \\
\text{low} & \leftrightarrow & \text{mod}_6 & \leftrightarrow & \phi(\text{mod}_6)(t) = t^{\alpha_6}, & t \in [0, 1] & \\
\text{very low} & \leftrightarrow & \text{mod}_7 & \leftrightarrow & \phi(\text{mod}_7)(t) = t^{\alpha_7}, & t \in [0, 1] &
\end{array}$$

where for suitably chosen values the exponents order as

$$\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 = 1 > \alpha_5 > \alpha_6 > \alpha_7 > 0 \,. \tag{7.6}$$

Note also the typical interpretations

$$\text{not high} \leftrightarrow 1 - \text{mod}_2 \leftrightarrow 1 - t^{\alpha_2} \,, \tag{7.7}$$

etc.

Next, define the *basic matching attributes* match (A_i) as follows:
Pick a copula $\text{cop}_i : [0, 1]^2 \rightarrow [0, 1]$ that appears most appropriate and define

$$\phi(\text{match}(A_i))(z_{i,1}, z_{i,2}) = \text{cop}_i(\phi(A_i)(z_{i,1}),\ \phi(A_i)(z_{i,2})) \tag{7.8}$$

for all $z_{ij} \in D_{A_i}$, $j = 1, 2; i = 1, .., \text{m}$. The purpose here is to describe the *joint* distributional value of $z_{i,1}$ and $z_{i,2}$, not just the "distance" (actual, statistical, or possibilistic), and it appears equation (7.8) captures this appropriately.

For example, returning to equation (7.1), note that (omitting subscript commas when unambiguous) for all $z_{11}, z_{12} \in D_1 \subsetneq \mathbf{R}^2$, by use of Schweizer's inequality and eigenvalue properties, it easily follows that equation (7.2) yields for $\mathrm{cop}_1 = \mathrm{prod}$, for example,

$$\begin{aligned} \phi(\mathrm{match}(\mathrm{geo}))(z_{11}, z_{12}) &\overset{d}{=} \phi(\mathrm{geo}(z_{11})) \cdot \phi(\mathrm{geo})(z_{12})) \\ &= \mathrm{rel}(z_1, z_2) \cdot \mathrm{abs}(z_1, z_2) , \end{aligned} \tag{7.9}$$

where

$$\begin{aligned} \mathrm{rel}(z_1, z_2) &= \lambda_{11}^2 e^{-\lambda_{12} \cdot (z_{11}-z_{12})^T \Lambda_1^{-1} (z_{11}-z_{12})} \\ \mathrm{abs}(z_1, z_2) &= e^{-2\lambda_{12} \cdot (z_{11}-z_{12})^T \Lambda_1^{-1} (z_{11}-z_{12})} \end{aligned} \tag{7.10}$$

$$\geq \exp\left[-2\lambda_{12} \left(\frac{\| z_{12} - \mu_1 \|^2}{\mathrm{maxeig}(\Lambda_1)} - \frac{\epsilon_o \cdot \| z_{12} - \mu_1 \|}{\mathrm{mineig}(\Lambda_1)} \right) \right] \tag{7.11}$$

for all

$$\| z_{11} - z_{12} \| \leq \epsilon_o . \tag{7.12}$$

Hence

$$\lim_{\left(\substack{\| z_{12} - \mu_1 \| \to +\infty , \\ \| z_{11} - z_{12} \| \leq \epsilon_o} \right)} (\mathrm{abs}(z_1, z_2)) \downarrow 0 , \tag{7.13}$$

all

$$\| z_{12} - \mu_1 \| > (\mathrm{maxeig}(\Lambda_1)/\mathrm{mineig}(\Lambda_1) \cdot \epsilon_o . \tag{7.14}$$

Thus,

$$\lim_{\left(\substack{\| z_{12} - \mu_1 \| \to +\infty , \\ \| z_{11} - z_{12} \| \leq \epsilon_o} \right)} \phi(\mathrm{match}(\mathrm{geo})(z_{11}, z_{12})) = 0 , \tag{7.15}$$

showing z_{11} and z_{12} being close to each other is not enough for matching relative to the parent attribute.

Note the following two important opposite limiting deterministic subcases:

(i) $\Lambda_1 \to 0$, i.e., maxeig $(\Lambda_1) \to 0$. This is equivalent to

$$\phi(\mathrm{geo}_2)(z) = \delta_{z,\mu_1} \quad \text{(Kronecker delta)} , \tag{7.16}$$

for all $z \epsilon \mathbf{R}^2$, in which case, for any copula

$$\phi(\text{match}(\text{geo}_2))(z_1, z_2) = \delta_{z_1}, \mu_1 \cdot \delta_{z_2}, \mu_1$$

$$= \begin{cases} 1, \text{if } z_1 = z_2 = \mu_1 \\ 0, \text{if } z_1, z_2 \text{ are otherwise} \end{cases} \tag{7.17}$$

for all $z_1, z_2 \epsilon \mathbf{R}^2$.

This case is not satisfactory, since the above indicates that one *knows absolutely* that, from a prior viewpoint, the target was at μ and nowhere else!

(ii) $\Lambda_1 \rightarrow \infty$, i.e., mineig $(\Lambda_1) \rightarrow \infty$.

This is equivalent to, regardless of copula chosen,

$$\phi(\text{match}(\text{geo}_2))(z_1, z_2) = \text{cop}_1(1, 1) = 1 , \tag{7.18}$$

meaning that for all z_1,z_2 in the truncated region of $\mathbf{R}^2$, the basic matching level is indifferent, i.e., all are equally likely!

Similarly, for A_2 in equation (7.3) with membership function described in equation (7.4), while BR(own) and BL(ack) are colors "close to" each other, their relations to A_2 are both seen to be small and hence BR and BL should not have a good match. Indeed, the maximal possible copula min shows

$$\phi(\text{match}(A_2))(BR, BL) = \min(0.2, 0.1) = 0.1 , \tag{7.19}$$

etc.

In turn, the *conditional basic matching attribute membership functions* are of the form, for any observed data set

$$y \overset{d}{=} (y_{i,1}, y_{i,2})_{i=1,\dots,m} \in Y \overset{d}{=} \underset{i=1}{\overset{m}{\times}} (D_{A_i} \times D_{A_i}) , \tag{7.20}$$

$y_{ij} \in D_{A_i}$, $j = 1, 2;\ i = 1, \dots, m$.

$$\phi(\text{match}(A_i))(z_i | y_i)$$

$$= \phi(\text{match}(A_i))(z_{i1}, z_{i2} | y_{i1}, y_{i2}) \tag{7.21}$$

$$= \text{cop}_i(\phi(A_i)(z_{i1} | y_{i1}), \phi(A_i)(z_{i2} | y_{i2})) .$$

If one knows the antecedent function g in equation (6.7)—not often so in practice—$\phi(A_i)(z_{ij} | y_{ij})$ can be obtained as a conditional fuzzy-set membership function. On the other hand, the function may be obtainable directly from experts in tabular form $\phi(A_i)(z_{ij} | y_{ij})$ versus z_{ij} and y_{ij}. For simplicity, one could assume

$$\phi(A_i)(z_{ij}|y_{ij}) = \psi_i(\phi(A_i)(z_{ij});y_{ij}) , \tag{7.22}$$

for some function ψ_i of two variables obtained empirically. (If $\psi_i(t,s) = \mathrm{cop}_i(t, \ _i(s))/ \ _i(s)$, essentially equation (6.1) is obtained back.)

8. DEVELOPMENT OF MODIFIED PACT: INFERENCE RULES AND ERRORS

Next, suppose a collection of inference rules has been culled from experts, connecting correlation variable x indirectly with observed data y, through auxiliary variables z: Specifically, let the collection of corresponding conditional membership functions be denoted as, for any possible true values

$$z \overset{d}{=} (z_{i1}, z_{i2})_{i=1,..,m} \in Z \overset{d}{=} \mathop{\times}_{i=1}^{m} (D_{A_i} \times D_{A_i}), \tag{8.1}$$

$$z_{ij} \in D_{A_i}, \quad j = 1, 2; \quad i = 1, .., m,$$

$\{\text{rule}_k(x, z) : k = 1, .., r\}$, is defined by

$$\text{rule}_k \overset{d}{=} \text{rule}_k(x|z) = (\text{cons}_k(x)|\text{ant}_k(z)) \tag{8.2}$$

$$\text{cons}_k(x) \overset{d}{=} \phi(\text{mod}_{j_{k,o}})(\phi(A_o)(x)), \quad x \in [0, 1] \tag{8.3}$$

and

$$\text{ant}_k(z) = \text{comb}_k(\&, or, not)(m_k(z)) , \tag{8.4}$$

where

$$m_k(z) = (\phi(\text{mod}_{j_{k,i}})[\phi(\text{match}(A_i))(z_i)])_{i \in J_k} , \tag{8.5}$$

$$z_i = (z_{i1}, z_{i2}), \quad i = 1, .., m . \tag{8.6}$$

Here, comb(&, *or*, *not*) is some well-defined logical combination of a copula $\text{cop}_{(k)}[0, 1]^2 \to [0, 1]$, cocopula $\text{cocop}_{(k)}[0, 1]^2 \to [0, 1]$, and negation $1 - (\ ) : [0, 1] \to [0, 1]$) applied to the modifications $\text{mod}_{j_{k,i}}$ of matches. Also, depending on $k, k = 1, ., r$,

$$j_{k,i} \in \{1, .., 7\}, \quad i = 0, 1, .., m , \tag{8.7}$$

and index set J_k of attributes satisfies

$$\emptyset \neq J_k \subseteq \{1, .., m\} . \tag{8.8}$$

For example, one might have the following three rules, given also in symbolic exponential forms:

$$
\begin{aligned}
\text{rule}_1 \leftrightarrow\ & \text{"If } A_1 \text{ matches very high, but } A_2 \text{ matches low,} \\
& \text{then correlation level is moderately high"} \\
\leftrightarrow\ & (A_o^{\alpha_3} | A_1^{\alpha_1} \ \& \ A_2^{\alpha_6}) \ ,
\end{aligned}
\tag{8.9}
$$

where

$$J_1 = \{1, 2\}; \quad j_{10} = 3; \quad j_{11} = 1; \quad j_{12} = 6 \ ; \tag{8.10}$$

$$
\begin{aligned}
\text{rule}_2 \leftrightarrow\ & \text{"If } A_1 \text{ does not match high, or if } A_3 \\
& \text{matches moderately high then correlation} \\
& \text{level is low medium"} \\
\leftrightarrow\ & (A_o^{\alpha_5} | (1 - A_1^{\alpha_1} \ or \ A_3^{\alpha_3}) \ ,
\end{aligned}
\tag{8.11}
$$

where

$$J_2 = \{1, 3\}; \quad j_{20} = 5; \quad j_{21} = 1; \quad j_{23} = 3 \ ; \tag{8.12}$$

$$
\begin{aligned}
\text{rule}_3 \leftrightarrow\ & \text{"If } A_1 \text{ matches high and } A_2 \text{ matches} \\
& \text{mediumly and } A_3 \text{ matches very low,} \\
& \text{then correlation level is low"} \\
\leftrightarrow\ & (A_o^{\alpha_6} | A_1^{\alpha_2} \ \& \ A_2^{\alpha_4} \ \& \ A_3^{\alpha_7}) \ ,
\end{aligned}
\tag{8.13}
$$

where

$$J_3 = \{1, 2, 3\}; \quad j_{30} = 6; \quad j_{31} = 2; \quad j_{32} = 4; \quad j_{33} = 7 \ . \tag{8.14}$$

The k^{th} inference rule at values x and z states basically "if the logical combination of modifiers of matches $\text{ant}_k(z)$ holds, then, the correlation level $\text{cons}_k(x)$ holds" or equivalently "the degree of compatibility of (x,z) relative to the conditional form $(\text{cons}_k | \text{ant}_k)$.

Finally, with all of the above established, equation (6.7) can be utilized as follows: For each inference rule k, $k = 1, \ldots, r$, let

$$f(x) \overset{d}{=} \text{cons}_k(x) \ , \tag{8.15}$$

as in equation (8.3);

$$g(y, z) \overset{d}{=} \text{ant}_k(z) \ , \tag{8.16}$$

as in equations (8.4) through (8.6), and let

$$
\begin{aligned}
(h(z)|g(y)) &\overset{d}{=} \text{ant}_k(z|y) \overset{d}{=} (\text{ant}_k(z)|y) \\
&\overset{d}{=} \text{ant}_k(z)\,(\text{with z replaced by } z|y) \\
&\overset{d}{=} \text{comb}_k(\&, or, not)(m_k(z|y)) \ ,
\end{aligned}
\tag{8.17}
$$

where

$$
m_k(z|y) = (\phi(\text{mod}_{j_{k,i}})[\phi(\text{match}(A_i)(z_i|y_i))])_{i \in J_k}
\tag{8.18}
$$

a *conditional error form* for the k^{th} inference rule antecedent.

Substituting equations (8.15) through (8.18) into equation (6.7), using equation (8.2), yields for all $x \in [0, 1]$, the k^{th} *posterior form*

$$
\text{cons}_k(x|y) \overset{d}{=} (\text{cons}_k(x)|y) = \max_{z \in D} (\text{rule}_k(x|z) \cdot \text{ant}_k(z|y)) \ ,
\tag{8.19}
$$

where $z \in D$ means $(z_{i1}, z_{i2} \in D_{A_i} \ \ i = 1, .., \text{m})$.

If the simplifying form in equation (7.22) is assumed, then equation (8.19) also simplifies, by substituting

$$
t_{ij} \overset{d}{=} \phi(A_i)(z_{ij}), \quad j = 1, 2; \ i = 1, \ldots, \text{m} \ ,
\tag{8.20}
$$

and letting

$$
\underline{t} \overset{d}{=} (t_{i1}, t_{i2})_{i=1,\ldots,\text{m}} \ .
\tag{8.21}
$$

One obtains a reasonable upper-bound approximation (since range $(\phi(A_i)) \subset [0, 1]$ in general):

$$
\begin{aligned}
\text{cons}_k(x|y) &\leq \text{cons}_k(x|y)_o \\
&\overset{d}{=} \max_{\underline{t} \in [0,1]^{2m}} ((\text{cons}_k(x)|\alpha_k(\underline{t})) \cdot \alpha_k(\psi_k(\underline{t}, y))) \ ,
\end{aligned}
\tag{8.22}
$$

where

$$
\alpha_k(\underline{t}) = \text{comb}_k(\&, or, not)(\eta_k(\underline{t})) \ ,
\tag{8.23}
$$

$$
\eta_k = (\phi(\text{mod}_{j_{k,i}})[\text{cop}_i(t_{i1}, t_{i2})])_{i \in J_k} \ .
\tag{8.24}
$$

$$
\alpha_k(\psi_k(\underline{t}, y)) = \text{comb}_k(\&, or, not)(\psi_k(\underline{t}, y)) \ ,
\tag{8.25}
$$

$$
\psi_k(\underline{t}, y) = (\phi(\text{mod}_{j_{ki}})[\text{cop}_i(\psi_i(t_{i1}, y_{i1}), \psi_i(t_{i2}, y_{i2}))])_{i \in J_k}
\tag{8.26}
$$

In turn, the r posterior forms $cons_k(X|y)$, $k = 1,..,r$, must be combined into one overall posterior.

By the *diagonalization technique* (reference 14), one can naturally combine the separate posteriors in equation 8.19 as the solution, $cons(X|y)$, of the equation

$$\mathrm{cop}(\mathrm{cons}(x|y), \mathrm{cons}(\mathrm{diag}|y)) = \mathrm{cons}_o(x|y) \ , \tag{8.27}$$

where

$$\begin{aligned} &\mathrm{cons}_o(x|y) \overset{d}{=} \mathrm{cop}((\mathrm{cons}_k(x)|y))_{k=1,\ldots,r}) \ , \\ &\mathrm{cons}(\mathrm{diag}|y) = \max_{x\epsilon[0,1]} (\mathrm{cons}_o(x|y)) \ . \end{aligned} \tag{8.28}$$

9. SUMMARY OF INPUTS AND COMPUTATIONS

For both the full-modified PACT form in equation (8.19) and the shortened upper-bound form in equation (8.22), the following basic input steps are required:

1. Determine the master list of auxiliary attributes $A_1, \ldots, A_m$ and obtain observed/updated/reported data $y_{ij} \in D_{A_i}$ for target j relative to A_i, $j = 1, 2$, $i = 1, \ldots, m$.

2. Obtain the table of appropriate values of exponents a_ℓ, $\ell = 1, \ldots, 7$ in equations (7.5) and (7.7) for interpreting modifiers mod_ℓ, $\ell = 1, \ldots, 7$.

3. Provide the master list of inference rules $\text{rule}_1, \ldots, \text{rule}_r$ from experts in syntactic form. The syntactic structure is given as

$$\text{rule}_k \leftrightarrow (\underbrace{a_{j_{k,o}}}_{\substack{\textit{conseq.}\\ \textit{exponent}}};\ \underbrace{\text{comb}_k(\&, or, not)((a_{j_{k,i}})_{i \in J_k})}_{\substack{\textit{antecedent}\\ \textit{exponent combination}}})$$

for some $\emptyset \neq J_k \subseteq \{1, \ldots, m\}, k = 1, \ldots, r$.

If, in particular, $\text{comb}_k(\&, or, not)$ is strictly $\text{cop}_{(k)}$, then one can write

$$\text{rule}_k \leftrightarrow (a_{j_{k,o}}; (a_{j_{k,i}})_{i \in J_k}).$$

4. Determine for each attribute A_i, the membership function $\phi(A_i) : D_{A_i} \to [\,0, 1]$, $i = 0, 1, 2, \ldots, m$.

5. Determine the appropriate forms for: copula cop_i, corresponding to match (A_i), i=1, ..., m; copula cop_k and cocopula cocop_k for rule_k, $k = 1, \ldots, r$, and copula cop for use in combining the PACT outputs for each inference rule.

For only the full-modified PACT form in equation (8.19), the following input is required:

6. Determine, directly via experts, for each attribute A_i, the conditional membership function $\phi(A_i)(\cdot\,|y_i) : D_{A_i} \to [\,0, 1]$, for all $y_i \in D_{A_i}$, $i = 1, \ldots, m$

For only the upper-bound form for PACT in equation (8.22), the following input is needed:

7. Determine, directly via experts, for each attribute A_i, the function ψ_i relating $\phi(A_i)(z_{ij})$ and y_{ij} in the conditional form given in equation (7.22), i=1, ..., m. Thus, utilizing $\phi(A_i)$ and observation y_{ij}, one wishes to obtain the transfer relation

$$\phi(A_i)(z_{ij}) \to \psi_i(\phi(A_1)(z_{ij}), y_{ij})$$

when y_{ij} is observed, j=1,2; i=1, ..., m.

For the full-modified PACT form in equation (8.19), table 9-1 provides a summary of the required computations, with figure 9-1 giving the corresponding full flowcharts.

Table 9-1. Summary of computations determining PACT relative to the k^{th} given inference rule, in its full-modified form.

Compute for all $z_{i1}, z_{i2} \epsilon D_{A_i}$,
and all $x \epsilon [0, 1]$, with $y_{i1}, y_{i2} \epsilon D_{A_i}$ given, $i = 1, .., m$:

$$\phi(\text{match}(A_i))(z_i) = \text{cop}_i(\phi(A_i)(z_{i1}), \phi(A_i)(z_{i2})) \ , \tag{7.8}$$

$$m_k(z) = (\phi(\text{mod}_{j_{k,i}})[\phi(\text{match}(A_i))(z_i)])_{i \epsilon J_k} \ , \tag{8.5}$$

$$\text{ant}_k(z) = \text{comb}_k(\&, or, not)(m_k(z)) \ , \tag{8.4}$$

$$\text{cons}_k(x) = \phi(\text{mod}_{j_{k,o}})(\phi(A_o)(x)) \ , \tag{8.3}$$

$$\text{rule}_k(x|z) = (\text{cons}_k(x)|\text{ant}_k(z)) \tag{8.2}$$

$$= \text{cop}_{(k)}(\text{cons}_k(x), \text{ant}_k(z))/\text{ant}_k(z) \ , \tag{6.1}$$

provided $\text{ant}_k(z) > 0$.

$$\phi(\text{match}(A_i))(z_i|y_i) = \text{cop}_i(\phi(A_i)(z_{i1}|y_{i1}), \phi(A_i)(z_{i2}|y_{i2})) \ , \tag{7.21}$$

$$m_k(z|y) = (\phi(\text{mod}_{j_{k,i}})[\phi(\text{match}(A_i)(z_i|y_i)])_{i \epsilon J_k} \tag{8.18}$$

$$\text{ant}_k(z|y) = \text{comb}_k(\&, or, not)(m_k(z|y)) \ , \tag{8.17}$$

$$\text{cons}_k(x|y) = \max_{z \epsilon D}(\text{rule}_k(x|z) \cdot \text{ant}_k(z|y)) \ . \tag{8.19}$$

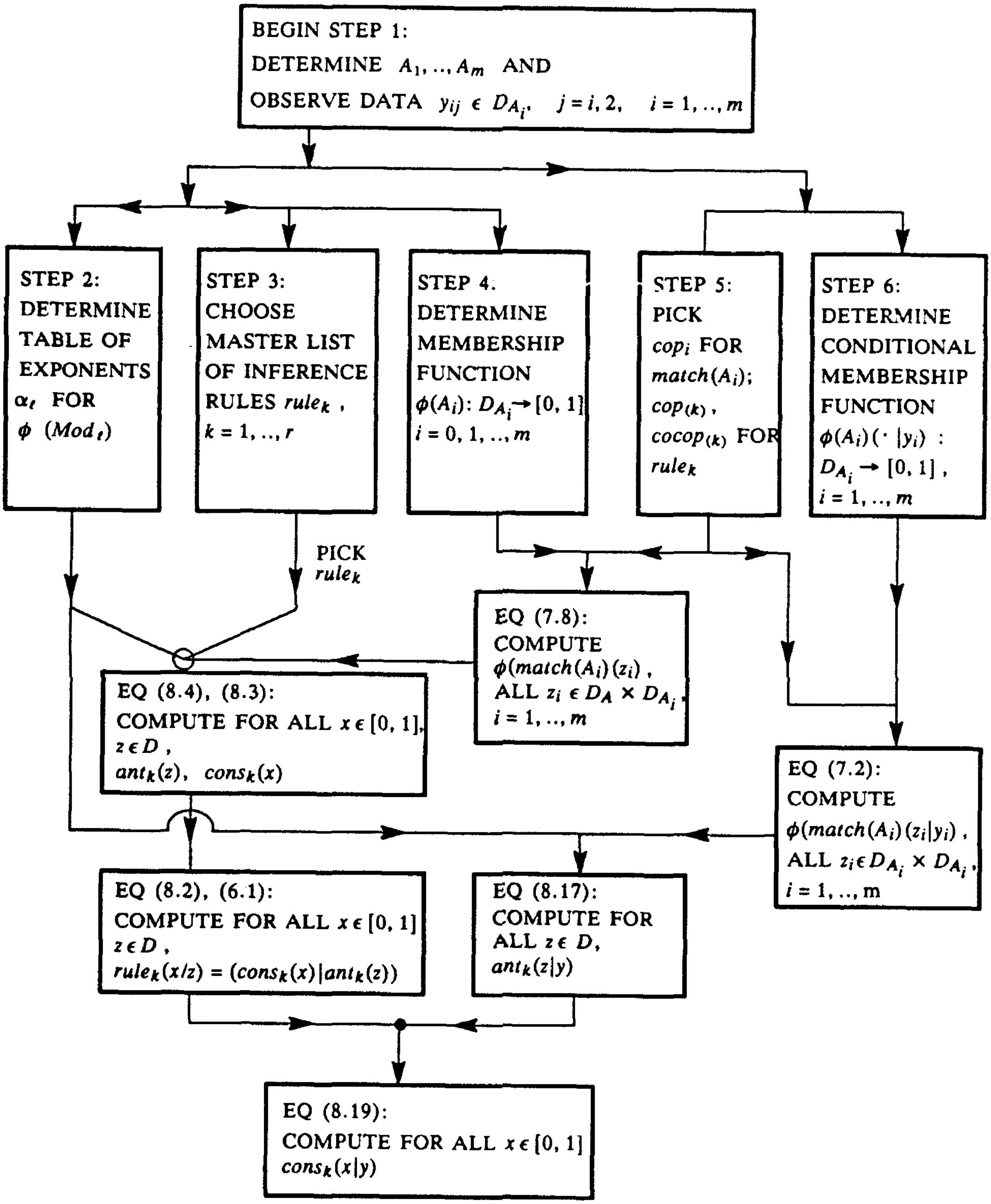


Figure 9-1. Flowchart of computations for determining full-modified PACT form relative to the k^{th} inference rule.

For the shortened upper-bound form for PACT in equation (8.22), table 9-2 provides a summary of the required computations, with figure 9-2 giving the corresponding full flowcharts.

Table 9-2. Summary of computations determining the shortened upper-bound form for the k^{th} given inference rule.

Compute for all $t_{11}, t_{12}, x \epsilon [0, 1]$, $i = 1, .., m$:

$$n_k(\underline{t}) = (\phi(\text{mod}_{j_{k,i}})[\text{cop}_i(t_{i1}, t_{i2})])_{i \epsilon J_k}, \quad (8.24)$$

$$\alpha_k(\underline{t}) = \text{comb}_k(\&, or, not)(n_k(\underline{t})), \quad (8.23)$$

$$\text{cons}_k(x) = \phi(\text{mod}_{j_{k,o}})(\phi(A_o)(x)), \quad (8.3)$$

$$(\text{cons}_k(x) | \alpha_k(t)) = \text{cop}_{(k)}(\text{cons}_k(x), \alpha_k(t)) / \alpha_k(t), \quad (6.1)$$

provided $\alpha_k(t) > 0$,

$$\psi_k(\underline{t}, y) = (\phi(\text{mod}_{j_{ki}})[\text{cop}_i(\psi_i(t_{i1}, y_{i1}), \psi_i(t_{i2}, y_{i2}))])_{i \epsilon J_k}. \quad (8.26)$$

$$\alpha_k(\psi_k(\underline{t}, y)) = \text{comb}_k(\&, or, not)(\psi_k(\underline{t}, y)), \quad (8.25)$$

$$\text{cons}_k(x|y)_o = \max_{\underline{t} \epsilon [0,1]^{2m}} (\text{cons}_k(x) | \alpha_k(t)) \cdot \alpha_k(\psi_k(\underline{t}, y))) \quad (8.22)$$

For both the full-modified and shortened versions of PACT, table 9-3 provides a summary of the required computations needed to combine the separate outputs of PACT relative to each inference rule.

Table 9-3. Summary of computations determining the shortened upper-bound form for each inference rule.

Compute for all $x \epsilon [0, 1], y \epsilon D$ given (i.e., $y_{i,1}, y_{i,2} \epsilon D_{A_i}$ given, $i = 1, .., m$) :

$$\text{cons}_o(x|y) = \text{cop}((\text{cons}_k(x|y))_{k=1,...,r}), \quad (8.28)$$

$$\text{cons}(\text{diag}|y) = \max_{x \epsilon [0,1]} (\text{cons}_o(x|y)) : \quad (8.29)$$

Solve for $\text{cons}(x|y)$:

$$\text{cop}(\text{cons}(x|y), \text{cons}(\text{diag}|y)) = \text{cons}_o(x|y). \quad (8.27)$$

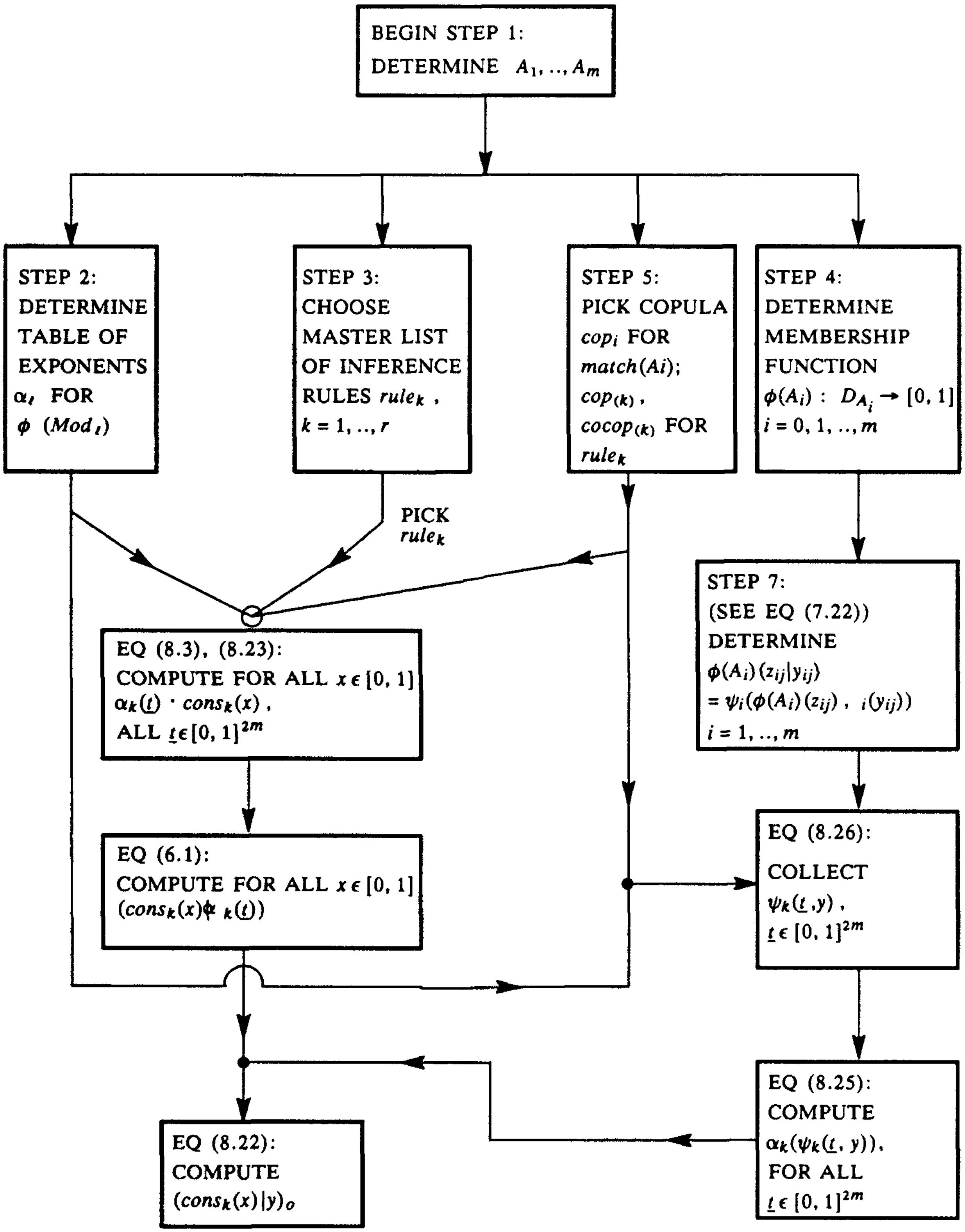


Figure 9-2. Flowchart of computations for determining shortened upper-bound version of PACT relative to the k^{th} inference rule.

10. AN ILLUSTRATIVE EXAMPLE

$m = 3$ auxiliary attributes: (10.1)

$A_1 = geo_2, \quad A_2 = color, \quad A_3 = radar_2$, (10.2)

Modifier exponents:
$\alpha_1 = 4, \quad \alpha_2 = 3, \quad \alpha_3 = 2, \quad \alpha_4 = 1, \quad \alpha_5 = 0.75, \quad \alpha_6 = 0.5, \quad \alpha_7 = 0.25$. (10.3)

$\text{cop}_i = \text{prod}, \quad \text{cocop}_i = \text{probsum}, \quad i = 1, 2, 3$;
$\text{cop} = \min, \quad \phi(not) = 1 - (\ )$. (10.4)

$r = 3$ inference rules: $\text{rule}_1, \text{rule}_2, \text{rule}_3$, as given in equations (8.9) through (8.14), with logical operators

$\text{cop}_{(1)} = \text{cop}_{(3)} = \text{prod}, \quad \text{cop}_{(2)} = \min$. (10.5)

For A_j, $D_{A_j} = \mathbf{R}^2$, $j = 1, 3$, with
$\mu_j \in \mathbf{R}^2, \Lambda_j$ 2 by 2 positive definite matrix ,
$\lambda_{j,1}, \lambda_{j,2} > 0$ all fixed ,
$\phi(A_j(z_j)) = \lambda_{ji} e^{-\lambda_{j2}(z_j-\mu)^T \Lambda_j^{-1}(z_j-\mu)}$, (10.6)

For all $z_j \in \mathbf{R}^2, j=1,3$;

For A_2, $D_{A_2} = \{YEL, OR, BL, WH, BR\}$ with
$\phi(A_2) : D_{A_2} \rightarrow [0, 1]$ given by
$\phi(A_2)(YEL) = 0.6, \quad \phi(A_2)(OR) = 0.5, \quad \phi(A_2)(BL) = 0.1$,
$\phi(A_2(WH) = 0.9, \quad \phi(A_2)(BR) = 0.2$. (10.7)

To determine ψ_i (equation 7.22) for $i=1,3$:

First, recall the following (see, for example, Anderson, reference 15, chapter 1):

Let random variables z, k_1 by 1, y, k_2 by 1, be jointly distributed $N_{k_1+k_2}(\mu, \Lambda)$, where

$$\mu = \begin{pmatrix} \mu_1 \\ \hline \mu_2 \end{pmatrix}, \quad \Lambda = \left(\begin{array}{c|c} \Lambda_1 & \Lambda_{12} \\ \hline \Lambda_{12}^T & \Lambda_2 \end{array}\right) , \tag{10.8}$$

μ, $k_1 + k_2$ by 1, with μ_i, k_i by 1 constant (mean), Λ, $k_1 + k_2$ by $k_1 + k_2$ positive definite with Λ_i, k_i by k_i positive definite and Λ_{12}, k_1 by k_2 constants, i=1,2. Then, the conditional random variable $(z|y)$ is distributed $N_{k_1}(E(z|y), Cov(z|y))$, where

$$E(z|y) = \mu_1 + \Lambda_{12} \cdot \Lambda_2^{-1} \cdot (y - \mu_2) \ , \tag{10.9}$$

$$Cov(z|y) = \Lambda_1 - \Lambda_{12}\Lambda_2^{-1}\Lambda_{12}^T \ . \tag{10.10}$$

Next, suppose the simple additive linear regression relation holds

$$y = z = w \ , \tag{10.11}$$

where z and w are statistically independent with z distributed $N_{k_o}(\mu_1, \Lambda_1)$ (Λ_1 pos. def.) and w distributed $N_{k_o}(\mu_3, \Lambda_3)$ (Λ_3 pos. def.).

Then $\binom{z}{y}$ is jointly distributed and equations (10.8) through (10.10) hold, where now

$$\Lambda_{12} = \Lambda_1, \quad \Lambda_2 = \Lambda_1 + \Lambda_3 \ , \tag{10.12}$$

$$\mu_2 = \mu_1 + \mu_3 \ , \tag{10.13}$$

whence

$$E(z|y) \ = \ \mu_1 + \Lambda_1(\Lambda_1 + \Lambda_3)^{-1}(y - \mu_1 - \mu_3) \ , \tag{10.14}$$

$$Cov(z|y) \ = \ \Lambda_1 - \Lambda_1(\Lambda_1 + \Lambda_3)^{-1}\Lambda_1 \ = \ (\Lambda_1^{-1} + \Lambda_3^{-1} \ )^{-1} \ . \tag{10.15}$$

It follows that for the normalized version of Gaussian distributions, letting $\lambda_{11} = \lambda_{12} = 1$ in equation 7.2, one has for its possibility/probability function counterpart

$$P(z|y) = e^{-\frac{1}{2}Q(z,y)}, \ z \in \mathbf{R}^2, \tag{10.16}$$

where by expanding

$$\begin{aligned} Q(z,y) &\overset{d}{=} (z - E(z|y))^T Cov^{-1}(z|y)(z - E(z|y)) \ = (z - \mu_1)^T \cdot (\Lambda_1^{-1} + \Lambda_3^{-1})^{-1} \cdot (z - \mu_1) \\ &\quad + \big((y - \mu_1 - \mu_3)^T \cdot R \cdot (y - \mu_1 - \mu_3)\big) - 2S \ , \end{aligned} \tag{10.17}$$

where

$$\begin{aligned} R &\overset{d}{=} (\Lambda_1 + \Lambda_3)^{-1}\Lambda_1(\Lambda_1^{-1} + \Lambda_3^{-1})\Lambda_1(\Lambda_1 + \Lambda_3)^{-1} \\ &= \Lambda_3^{-1} \cdot (\Lambda_1^{-1} + \Lambda_3^{-1})^{-1} \cdot (\Lambda_1^{-1} + \Lambda_3^{-1})(\Lambda_1^{-1} + \Lambda_3^{-1})^{-1} \cdot \Lambda_3^{-1} \\ &= \Lambda_3^{-1} \cdot (\Lambda_1^{-1} + \Lambda_3^{-1})^{-1} \cdot \Lambda_3^{-1} \ , \end{aligned} \tag{10.18}$$

and

$$
\begin{aligned}
S &\stackrel{d}{=} (y-\mu_1-\mu_3)^T \cdot (\Lambda_1+\Lambda_3)^{-1} \cdot \Lambda_1 \cdot (\Lambda_1^{-1}+\Lambda_3^{-1}) \cdot (z-\mu_1) \\
&= (y-\mu_1-\mu_3)^T \cdot (\Lambda_1+\Lambda_3)^{-1} \cdot (\Lambda_1+\Lambda_3) \cdot \Lambda_3^{-1} \cdot (z-\mu_1) \\
&= (y-\mu_1-\mu_3)^T \cdot \Lambda_3^{-1} \cdot (z-\mu_1) \\
&= (y-\mu_1-\mu_3)^T \cdot \Lambda_3^{-\frac{1}{2}} \cdot \theta(z) \cdot \gamma(z) \quad ,
\end{aligned}
\tag{10.19}
$$

where

$$
\theta(Z) \stackrel{d}{=} \frac{1}{\gamma(z)} \cdot \Lambda_3^{-\frac{1}{2}} \cdot (z-\mu_1) \ , \tag{10.20}
$$

$$
\begin{aligned}
\gamma(z) &\stackrel{d}{=} \| \Lambda_3^{-\frac{1}{2}} \cdot (z-\mu_1) \| \\
&= ((z-\mu_1)^T \Lambda_3^{-1} (z-\mu_1))^{\frac{1}{2}} \ .
\end{aligned}
\tag{10.21}
$$

Note that by its very definition,

$$
\| \theta(z) \| \equiv 1, \ \text{all } z \neq \mu_1 \ . \tag{10.22}
$$

Next, make the conjugate-like assumption that

$$
\Lambda_1 = \kappa \cdot \Lambda_3 \ , \tag{10.23}
$$

for some real positive constant κ to be determined.

Denote the prior function for z as

$$
P(z) = e^{-\frac{1}{2}(z-\mu_1)^T \Lambda_1^{-1}(z-\mu_1)}, \quad z \in \mathbf{R}^{k_o} \ , \tag{10.24}
$$

corresponding to the normalized version of z being distributed $N_{k_o}(\mu_1, \Lambda_1)$ as above. We wish to solve for $P(z|y)$ in terms of $P(z)$ and y— and hence, à la equation 7.22 – determine the most appropriate function ψ_i (and g_i):

Leaving $\theta(z)$ alone temporarily, clearly applying equation (10.23) to equation (10.21) yields

$$
\begin{aligned}
\gamma(z) &= \kappa^{\frac{1}{2}} \cdot ((z-\mu_1)^T \Lambda_1^{-1} (z-\mu_1))^{\frac{1}{2}} \\
&= \kappa^{\frac{1}{2}} \cdot (-2 \log P(z))^{\frac{1}{2}} \ .
\end{aligned}
\tag{10.25}
$$

Using equation (10.23), equation (10.18) becomes

$$
R \ = (\kappa/\kappa+1)) \cdot \Lambda_3^{-1} = (\kappa^2/(\kappa+1)) \cdot \Lambda_1^{-1} \tag{10.26}
$$

and

$$\Lambda_1^{-1} + \Lambda_3^{-1} = (1+\kappa) \bullet \Lambda_1^{-1} \; , \tag{10.27}$$

from which

$$(z-\mu_1)^T (\Lambda_1^{-1} + \Lambda_3^{-1})(z-\mu_1)$$

$$= (1+\kappa)(z-\mu_1)^T \Lambda_1^{-1}(z-\mu_1) \; ,$$

whence

$$e^{-\frac{1}{2}(z-\mu_1)^T(\Lambda_1^{-1}+\Lambda_3^{-1})(z-\mu_1)} = (P(z))^{1+\kappa} \; . \tag{10.28}$$

Next, replacing $\theta(z)$ by the approximating constant

$$\theta_o \overset{d}{=} \theta(z), \text{ with } z \text{ replaced by } E(y) = \mu_1 + \mu_3 \; , \tag{10.29}$$

yields

$$\theta_o = (1/(\mu_3^T \Lambda_3^{-1} \mu_3)^{\frac{1}{2}}) \cdot \Lambda_3^{-\frac{1}{2}} \cdot \mu_3 \in \mathbf{R}^{k_o} \tag{10.30}$$

Then, gathering together equations (10.25) through (10.30) shows that under the assumptions in equations (10.23) and (10.29), by substitution into equation (10.17), equation (10.16) becomes

$$\begin{aligned} P(z|y) &\approx \psi(P(z), y) \\ &\overset{d}{=} A(y) \cdot B(P(z)) \cdot C(P(z), y) \; , \end{aligned} \tag{10.31}$$

where

$$A(y) \overset{d}{=} e^{-\frac{1}{2}(\kappa/(\kappa+1))(y-\mu_1-\mu_3)^T \Lambda_3^{-1}(y-\mu_1-\mu_3)} \; , \tag{10.32}$$

$$B(P(z)) \overset{d}{=} P(z)^{1+\kappa} \; , \tag{10.33}$$

$$C(P(z), y) \overset{d}{=} e^{\{(y-\mu_1-\mu_3)^T \cdot \Lambda_3^{-\frac{1}{2}} \cdot \theta_o \cdot \kappa^{\frac{1}{2}} \cdot (-2 \log P(z))^{\frac{1}{2}}\}} \; . \tag{10.34}$$

<u>Main Application</u>: Let $k_o = 2$, $P(z) = \phi(A_j)(z_j)$, $j = 1, 3$.

11. SUMMARY AND CONCLUSIONS

This technical document is a followup to the basic development of the PACT algorithm for combining linguistic-based and probabilistic information in track-data association or correlation. While the same spirit of PACT is followed (see reference 1 for background), a number of modifications are made in order to reduce the nonsystematic or ad hoc aspects. Those include: modeling of implication in PACT's inference rules by using conditional fuzzy sets or their appropriate approximations, rather than by fuzzification of material implication, as in the previous PACT form; utilization of one inference rule at a time to obtain the basic PACT posterior form for correlation, followed by a combining procedure (diagonalization [reference 14]), making use, when feasible, of a fuzzy-set extension of the author's previously established conditional-event algebra (reference 13); choosing copulas (rather than possible noncopula t-norms as in the previous version of PACT/and cocopulas for all logical connectors, based upon random-set considerations, outlined in Chapter 5).

Finally, a new shortened upper-bound approximation to the basic PACT output is exhibited, replacing the iterated disjunction operation over all auxiliary attribute domains by the more standardized domain of the unit interval, appropriately replicated. Based on the summary of inputs and computations (Chapter 9) for both the full-modified version of PACT and the shortened upper-bound one, it follows readily by inspection (especially of tables 9-1 and 9-2) that

$$\frac{\text{running time for shortened version of PACT}}{\text{running time for full modified version of PACT}} \approx \left(\frac{(n+1)^m}{\prod_{i=1}^{m} card(D_{A_i})}\right)^2, \tag{11.1}$$

where, as usual, m is the number of attributes, $card(D_{A_i})$ (assuming all attribute domains are finite or made finite) is the cardinality of D_{A_i}, $i = 1, .., m$, and, finally, n+1 is the number of elements in $\{0, \frac{1}{n}, \frac{2}{n}, .., \frac{n-1}{n}, 1\}$, the n^{th} discretized version of unit interval $[0,1]$.

Naturally, a tradeoff exists between the running time of the shortened version of PACT for the n^{th} discretized version of $[0,1]$ versus implementational fidelity versus accuracy relative to the full-modified version of PACT. Future numerical experiments will hopefully address this issue. Such experiments also are planned for testing the robustness and improvement in accuracy for the new versions of PACT relative to use—or lack of use—of linguistic-based attributes, for fixed geolocation and other statistical attribute information.

12. REFERENCES

1. Goodman, I. R. 1986. "PACT: An Approach to Combining Linguistic-Based and Probabilistic Information in Correlation and Tracking." NOSC TD 878 (March). Naval Ocean Systems Center, San Diego, CA.

2. Goodman, I. R., H. L. Wiener, and W. W. Willman. 1980. "Naval Ocean-Surveillance Correlation Handbook, 1979." NRL Report 8402. Naval Research Laboratory, Washington, DC. September 17, 1980. (Unclassified, but limited distribution)

3. Wiener, H. L., W. W. Willman, I. R. Goodman, and J. H. Kullback, "Naval Ocean Surveillance Correlation Handbook, 1978." NRL Report 8340. Naval Research Laboratory, Washington, DC. October 31, 1979. (Unclassified, but limited distribution)

4. Larsen, R. W., and R. A. Dillard. 1989. "A Fuzzy Logic Decision Support Tool." NOSC TR 1286 (October). Naval Ocean Systems Center, San Diego, CA.

5. Eisen, M. 1969. *Introduction to Mathematical Probability Theory*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

6. Dubois, D., and H. Prade. 1980. *Fuzzy Sets and Systems*. Academic Press.

7. Goodman, I. R., and H. T. Nguyen. 1985. *Uncertainty Models for Knowledge-Based Systems*. North-Holland Press, Amsterdam.

8. Goodman, I. R. 1980. "Identification of Fuzzy Sets With a Class of Canonically Induced Random Sets." *Proceedings of the 19th IEEE Conference on Decision and Control*, pp. 352-357.

9. Goodman, I. R. 1984. "Some New Results Concerning Random Sets and Fuzzy Sets." *Inf Sci* 34, pp. 93-113.

10. Goodman, I. R. "Algebraic and Probabilistic Bases for Fuzzy Sets and the Development of Fuzzy Conditioning," in the book, *Conditional Logic in Expert Systems* (I. R. Goodman, M. M. Gupta, H. T. Nguyen, G. S. Rogers, Ed.), North-Holland Press, Amsterdam (1991), pp. 1-69.

11. Schweizer, B., and A. Sklar. 1983. *Probabilistic Metric Spaces*. North-Holland Press, NY.

12. Goodman, I. R. 1990. "Three-Valued Logics and Conditional Event Algebras." *Proceedings of the First International Symposium on Uncertainty Models and Applications*, University of Maryland, College Park, Maryland, December 1990, pp. 31-37.

13. Goodman, I. R., H. T. Nguyen, and E. A. Walker. 1991. *Conditional Inference and Logic for Intelligent Systems: A Theory of Measure-Free Conditioning*. North-Holland Press, Amsterdam.

14. Goodman, I. R. 1989. "Applications of a Conditional Event Algebra to Data Fusion: Part 2." *Proceedings of the Third Tri-Service Data Fusion Symposium*. Applied Physics Laboratory, Johns Hopkins University, Laurel, Maryland, May 1989, pp. 181-193.

15. Anderson, T. W. 1984. *Introduction to Multivariate Analysis*. Second Edition, John Wiley & Sons, NY.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1992	Final: FY 90 – FY 91

4. TITLE AND SUBTITLE
A REVISED APPROACH TO COMBINING LINGUISTIC AND PROBABILISTIC INFORMATION IN CORRELATION

5. FUNDING NUMBERS
PE: 0602936N
PR: RV36121
WU: DN300036

6. AUTHOR(S)
Dr. I. R. Goodman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Ocean Systems Center
San Diego, CA 92152-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NOSC TR 1386

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Chief of Naval Research
OCNR
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

With the advent of the field of Artificial Intelligence (AI), new insights have been gained into information handling in general, and target data association, in particular. It is now clear that the traditional use of only stochastic geolocation information (i.e., vector positions, velocities, etc.) is not really adequate to describe the full data association problem. Rather, the situation can be markedly improved by using previously disgarded information in the form of linguistic descriptions or narratives, such as "probably very long," "appears to be flying a dark-blue flag," "is rather oblong in appearance," etc. However, at present, despite the inroads made with AI techniques, there is no adequate comprehensive theory covering the complete modeling of such information in conjunction with the classical measurement information of geolocation.

A previous attempt at addressing the above issue was formalized through the development of the PACT (Possibilistic Approach to Correlation and Tracking) algorithm developed at NOSC. The novelty of the algorithm was its explicit use of both linguistic-based evidence and probabilistic information through a common structure of a possibilistic/fuzzy-set model. The algorithm is essentially a generalization of a conditional form of the well-known total probability expansion theorem, an alternative to Bayes' theorem, when priors are not readily determinated; and a number of auxiliary attributes must be utilized to connect the parameter of interest—here, data association or correlation—with the observed data. Specifically, PACT consists of a disjunction over all attribute arguments of a conjunction of conditional fuzzy-set membership functions, one representing inference rules connecting data association with combinations of matching of attributes, the other representing conditional errors. A typical inference rule is the form "if geolocation values match moderately, but visual descriptions match highly, then the correlation level between the targets should be very high."

While numerical experiments indicated PACT improving upon the traditional approach of using only geolocational information, serious questions have arisen concerning the ad hoc nature of the choice of logical operators, the modeling of the conditional fuzzy-set membership functions, and last—but, not least—the overall running times required to implement the algorithm.

The current work seeks to remedy the above difficulties of PACT by re-examining the mathematical and structural foundations. The consequence of this is twofold:

1. Conditional fuzzy-set membership functions should *not* be identified as fuzzy-set extensions of the usual material-implication operator applied to appropriate argument membership functions, but instead as natural extensions of the author's newly developed conditional event algebra. This is presented in detail in the paper "Algebraic and probabilistic bases for fuzzy sets and the development of fuzzy conditioning" in the book, *Conditional Logic in Expert Systems* (I. R. Goodman et al., Ed., North-Holland Press, 1991). Roughly speaking, conditional events (a|b), (c|d), etc., are the "events inside the conditional probability evaluations p(a|b), p(c|d), etc. These can be combined logically prior to any probability evaluations p to yield again (conditional) events. Conditional fuzzy sets are the fuzzy-set extensions of these entities. Consequently, inference rules can now be put on a firmer mathematical basis relative to their interpretation and evaluations. On the other hand, for purposes of expediency, at times it may be more desirable to approximate such conditional fuzzy sets by simpler quantities.

2. A great reduction in computation time for implementing PACT, due to the disjunction operation reiterating upon attribute domains mentioned earlier, can be effected through an appropriate substitution and approximation technique. In effect, the disjunction can now be restricted, without loss of generality, in an upper-bound sense, to simplify replications of the more standardized unit interval.

14. SUBJECT TERMS

correlation	fuzzy sets
data association	possibility functions
knowledge-based systems	multivalues logic

15. NUMBER OF PAGES
45

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

UNCLASSIFIED

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE (include Area Code)	21c. OFFICE SYMBOL
Dr. I. R. Goodman	(619) 553–4014	Code 421

NSN 7540-01-280-5500

Standard form 298 (BACK)

UNCLASSIFIED

INITIAL DISTRIBUTION

Code 0012	Patent Counsel	(1)
Code 014	K. J. Campbell	(1)
Code 0141	A. Gordon	(1)
Code 0141	G. M. Dillard	(1)
Code 171	G. L. Ruptier	(1)
Code 027	V. Ware	(1)
Code 02712	Archive/Stock	(6)
Code 0274B	Library	(2)
Code 40	R. C. Kolb	(1)
Code 402	J. H. Maynard	(1)
Code 42	J. A. Salzmann	(1)
Code 421	D. L. Conwell	(1)
Code 421	M. J. Gherrity	(1)
Code 421	F. Kramer	(4)
Code 421	I. R. Goodman	(12)
Code 422	C. J. Funk	(1)
Code 422	M. C. Mudurian	(1)
Code 444	R. Larsen	(1)
Code 521	S. E. Apitz	(1)
Code 721	J. Aitkinhead	(1)
Code 721	J. L. Kaina	(1)
Code 741	S. C. McGirr	(1)
Code 782	M. J. Shensa	(1)

Defense Technical Information Center
Alexandria, VA 22304–6145 (4)

NCCOSC Washington Liaison Office
Washington, DC 20363–5100

Center for Naval Analyses
Alexandria, VA 22302–0268

Navy Acquisition, Research and Development
Information Center (NARDIC)
Washington, DC 20360–5000

GIDEP Operations Center
Corona, CA 91718–8000

NCCOSC Division Detachment
Warminster, PA 18974–5000